Quality Assurance and Artifacts in Clinical Spectroscopy

Roland Kreis, Ph.D.
Department for Clinical Research, Unit for MR Spectroscopy and Methodology
University Berne, Switzerland
roland.kreis@insel.ch http://www.cx.unibe.ch/dkf1/amsm/rkreis-0.htm

With the advent of reimbursability of MRS and of (semi-)automatic data acquisition, data processing and quantitation, the interest in clinical uses of MRS has increased and the threshold to really use MRS in the clinic has been drastically lowered for non-experts. Unfortunately, MRS is not like MRI, where many artifacts are easy to recognize and even laymen can judge the quality of an image reasonably well. In MRS, pitfalls are ubiquitous, but not necessarily eye-catching. Even worse, there is no agreement on what exactly is a good spectrum among experts. And the most common answer of experts to the question of how they judge the quality of their spectra, is: "it depends". And indeed it does depend on whether we are talking about single voxel or CSI data, about long or short TE spectra, about whether we consider near-normal spectra or widely pathologic spectra, whether they are spectra of the brain, or another organ, etc. Hence, I will try to give a personal view of the factors that should be considered when judging the quality and usability of a ^{1}H-MR spectrum[1]. I will first list commonly used quality criteria, then mention some major factors that affect quality and reliability of spectra, present examples of artifacts, and finally conclude with remarks on how to judge individual data sets and on how to present results.

Quality Criteria

Signal to noise ratio (SNR): The SNR is often defined in frequency domain (FD) as the height of the largest metabolite peak divided by the average (rms) amplitude of the noise in a signal-free part of the spectrum*, alternatively it can be based on signal area or, similarly, time domain amplitude vs. noise at the end of the FID. If frequency domain intensity is used, SNR depends inversely on linewidth. Low SNR can be remedied by choosing larger ROI's, increased scan time or optimized hardware (local receive coil, higher field). As illustrated in Fig. 1, FD-SNR also depends on the acquisition time, real time filtering and of course apodization. While high frequency noise gives the impression of low SNR, it may be largely irrelevant for the precision of peak fitting with restrictions or prior knowledge on line widths. SNR is often used to discard bad spectra. However, as SNR is directly reflected in the error estimates obtained from model fitting (so-called Cramer Rao minimum variance bounds, CRMVB[2]) and CRMVB are more directly linked to confidence limits, a rejection criterion based on CRMVB's for each metabolite seems more effective. SNR may be convenient when integrating well isolated peaks (e.g. CSI with long TE).

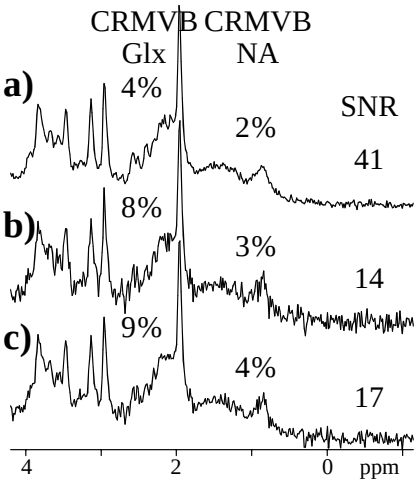

Fig. 1 *Identical spectrum with different noise level:* ***a)*** *original;* ***b)*** *noise added;* ***c)*** *same noise as b), but second half of FID replaced by zeroes. This changes visual appearance, but rather increases fitting errors (CRMVB).*

Linewidth/lineshape: Linewidth is usually defined independently of lineshape as the full width at half maximum peak height (FWHM) in FD. It determines the possible resolution of spectral features. If linewidths are estimated in model fitting they also influence the calculation of CRMVB's. It appears

* In LC-Model [32], it is not signal vs. noise, but vs. rms of the residues.

Abbreviations*:* ***Cho****: cholines;* ***Cr****: creatines,* ***CRMVB****: Cramer Rao minimum variance bounds;* ***CSF****: cerebro-spinal fluid;* ***CSI****: chemical shift imaging;* ***FD****: frequency domain;* ***FID****: free induction decay;* ***Glx****: glutamate + glutamine;* ***mI****; myo-inositol;* ***NA****: N-acetyl moieties;* ***NAAG****: N-acetylaspartylglutamate;* ***RF****: radio frequency;* ***rms****: root mean square;* ***ROI****: region of interest;* σ*;* ***SD****: standard deviation;* σ^2*: variance;* ***SNR****: signal to noise ratio;* ***SV****: single voxel;* ***TD****: time domain;* ***WS****: water suppression*

that linewidth is very critical for model fitting and bad resolution easily leads to meaningless results in short TE spectra. Rejection criteria based on FWHM are very useful for automatic screening of SV[3] and CSI[4] data. The effect of reduced resolution (and inherently lower intensity SNR) is illustrated in Fig. 2. It was recently found that reduced resolution and decreased SNR in combination with complicated baselines and possibly inaccurate fitting models can lead to systematic over- or under-estimation of low-concentration, but also prominent metabolites[5;6], which in part will not be reflected in increased CRMVB[7] (see below). Remedies for bad resolution include: better shimming (higher order shims, if available), smaller ROI size, moving ROI away from tissue interfaces.

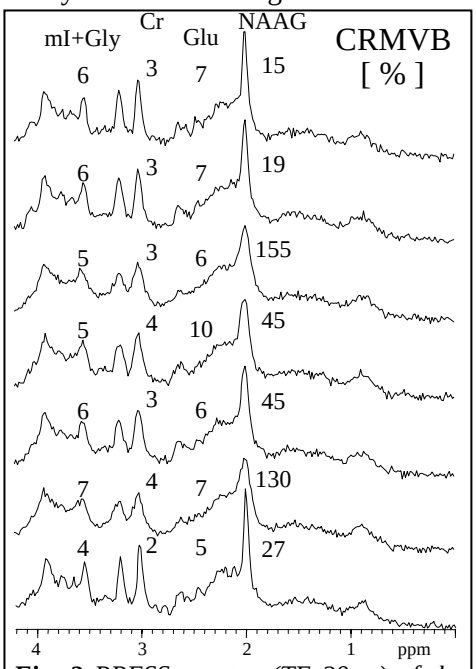

Fig. 2 *PRESS spectra (TE 20ms) of the same gray matter ROI with intentionally varied shim quality, demonstrating the covariation of linewidth and CRMVB, primarily for the minor contributors.*

Errors and variability: Quantitative results must always be given with error estimates which can be taken as a measure of spectral quality to a certain extent. The stochastic error for a single measurement is reflected in the Cramer Rao minimum variance bounds (CRMVB)[2;8] that include effects of the SNR and the inherent limits of fitting with a given model (including interdependence of fitting parameters and prior knowledge constraints). Hence, low CRMVB are good indicators of spectral quality. However, it is important to know that CRMVB are calculated under the premise that the fitting model is correct and complete[8]. Any systematic errors or artifacts are neglected and may lead to overoptimistic confidence limits. For the judgement of spectral quality in the context of a group of normal reference spectra, it is most useful to determine reproducibility. Reproducibility must be established locally and not be inferred from the literature. With regard to absolute quantitation in standard units, systematic errors are often substantial (e.g. inaccurate values for T_1, T_2, inaccuracies in reference measurements and calibrations, wrong baseline and lineshape models). The random errors, reflected in the reproducibility can be much smaller[9] than systematic errors and should not be taken as a measure for the latter. The quantitative method with best reproducibility does by no means guarantee its results to be closest to the true values. However, best reproducibility guarantees most sensitive detection of pathology.

Quality of data and fitted values can be judged based on error estimates from CRMVB, confidence limits[10], minimum χ^2, and/or general reproducibility at the local site. For CSI data, these parameters can be mapped leading to confidence images[10;11] or rejection masks[4].

Factors affecting spectral quality

Motion: In SV MRS, repeated small gross motions or local pulsatile motion (cardiac related CSF pulsation, respiration) normally lead to increased linewidths, possibly reduced peak areas (phase cancellation)[12] and decreased quality of water suppression. Single events of gross bulk motion will result in recording data from a wrong ROI. In the spectrum this may be evident by a doubling of all peaks (see Fig. 3). A post-acquisition correction is only possible, if all acquisitions were stored separately and if signals are present that allow for realignment and/or individual phasing. In CSI, motion leads to blurring[13], evidenced if substantial metabolite signals are found outside the head.

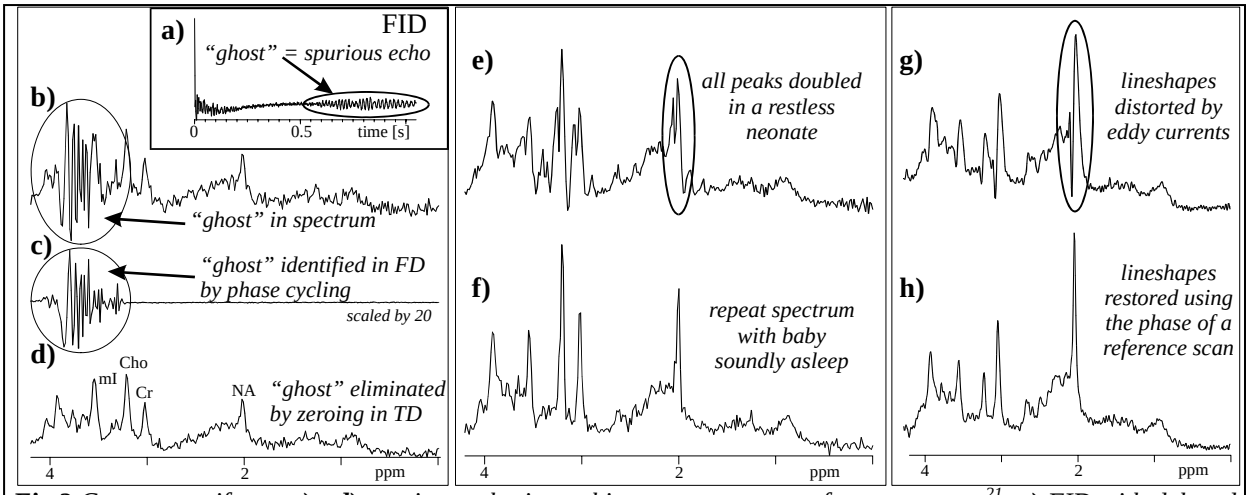

Fig.3 *Common artifacts:* ***a) - d)*** *spurious echo in a white matter spectrum from a neonate[21].* ***a)*** *FID with delayed spurious echo visible;* ***b)*** *spectrum incl. "ghost";* ***c)*** *"ghost" signal identified as a 2 pulse echo[26];* ***d)*** *ghost eliminated by truncation to first half of FID.* ***e) - f)*** *Effect of head movements in a neonatal spectrum from thalamus[21]:* ***e)*** *all peaks doubled due to movement;* ***f)*** *perfect lineshape without motion in repeat scan.* ***g) – h)*** *Effect of eddy currents:* ***g)*** *lineshapes distorted due to eddy currents in a short TE PRESS spectrum of occipital gray matter in a 14 year old;* ***h)*** *lineshape restored using the phase information from a water reference scan.*

ROI location: The correct placement of the ROI should always be checked to avoid wrong diagnoses because of typing errors or patient movements. Control of data header information, acquisition of a ROI image and the repetition of scout images after the recording of the MR spectrum are advisable.

Signal phase: The phase must be correct for visual inspection of the spectrum and good phasing may help to find the global minimum in fitting. Good phasing is not crucial for quantitative evaluation, if the phase is a variable in the fit, and the proper phase can usually be obtained from reference scans.

Water suppression (WS): Poor WS is not very problematic, if the shape of the residual water signal is well-behaved and accounted for in data fitting and if there are no appreciable vibrational sidebands. Often WS pulses saturate parts of the spectrum and can influence the determination of mI in short TE MRS.

Hardware problems: MRS may be first to be affected by failing hardware, as spectral quality and reproducibility depend strongly on optimal and constant hardware (shim, eddy currents, rf homogeneity and stability, amplifier linearity, external noise sources). Regular quality control[5;14-16] with phantom and volunteer scans is highly recommended. In particular, quantitation using the reciprocity principle relies heavily on constant receiver amplification and a linear RF amplifier. Hardware drifts during the scan can also lead to degraded spectral quality which can be remedied by inclusion of reference scans[17].

ROI shape: One should always have the real-world voxel profile[18;19] in mind, when judging spectra from focal lesions. It is not a sharp cube and one should remember that it is drastically altered for incorrect B_1 settings[18]. In CSI, the voxel size does not correspond to the ideal raw nominal size obtained by dividing field of view by number of phase steps.

Assignments: Never trust automatic peak assignments for grossly abnormal or low quality spectra. Often lipid signal is labeled as lactate or alanine. Automatisms can lead to Cho being labeled as Cr, etc.

Gallery of Artifacts

Artifact detection has not been automated and requires an experienced eye[1]. Some examples are presented in Fig. 3, others can be found under www.cx.unibe.ch/dkf1/amsm/MRS_artifacts.

- *Outer volume signal bleed:* ROI selection pulses are not infinitely selective and will always excite spins outside the targeted ROI as well (particularly within an ROI width around the ROI, and especially if B_1 was not adjusted properly). As the out of volume signal will normally be only a small fraction of the proper signal from the ROI, this will not be relevant, unless the surrounding tissue provides much stronger signals (lipids). Often, the out of volume signal is characterized by a different phase than the signal from within the ROI. RF-phase cycling, however, will not reduce the artifact. Spatial saturation bands can effectively prevent these signals.
- *Outer volume ghosts:* In 3-pulse spatial selection schemes (PRESS, STEAM) crusher gradient pulses are inserted to prevent refocusing of unwanted echoes or fids. Crusher gradient pulses work fine for homogeneous B_0. If appreciable local gradients are present at tissue interfaces, they can cancel the effect of some of the crusher gradients and, hence, lead to refocusing of unwanted echoes within the acquisition window. Refocusing hardly ever is at the expected echo maximum and this leads to the typical appearance of outer volume ghosts, depicted in Fig. 3. Enhanced phase cycling can reduce these effects and if individual scans are saved, can be used to identify these ghost signals[20].
- *Shifted echoes:* Ill-adjusted tweaking gradients result in shifted echoes. This leads to reduced peak areas and/or telltale lineshapes that urgently call for gradient readjustments.
- *Eddy currents:* Similarly, uncompensated eddy currents will lead to asymmetric lineshapes that are particularly striking for spectra with good resolution. Within certain limits, eddy current effects on lineshape can easily be corrected with the phase information from reference scans. On the down side, eddy current correction with inappropriate reference scans (recorded after patient moved, or containing substantial lipid signal in MRS outside the brain) introduces substantial sidebands. At longer TE, eddy currents also lead to signal drop in excess of the T_2 decay (possibly location-dependent).
- *Inadequate model:* Artifacts can often be detected in the fitting residuals. Unexpected features in the residuals could also be due to metabolites not included in the basis set. Automated detection of artifacts can be based on the size of χ^2 in comparison to what it should be for random noise.

General Considerations

To conclude, some potential guidelines to judge individual data sets:

Criteria for rejection of data: Based on above statements, the literature and an opinion poll among colleagues, criteria for rejection of spectra or individual metabolite values can be formulated (see text box).

Reject data if:

- **FWHM of metabolites > 0.07-0.1 ppm**
- **CRMVB > 30-50%**
- **unexplained features in residuals:** reject, if artifact, or expand model, if unexpected metabolite
- **peaks doubled**
- **lineshape strongly asymmetric** after eddy correction
- **outer volume ghosts present** (at least exclude metabolites overlaid with ghost)
- **out of phase signals present** (possibly just exclude Lac and Ala, if it is lipid, confined below 1.6 ppm)

Reproducibility: The achievable reproducibility has been published for several methods and brain regions and can serve as indication of what is feasible, even though the results scatter considerably. The overall reproducibility is determined (additive variances σ^2) by variability upon immediate repetition of a scan (σ_{rep}; closely related to CRMVB[21]), additional intra-individual variability upon re-examination of the same subject in a subsequent scan (σ_{intra})[†], and inter-individual variability (σ_{inter}). Ranges are given for NA, Cr, Cho, mI.

[†] $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2}$ and $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2 + \sigma_{inter}^2}$ correspond to the total variability upon reexamination of single subjects and to the overall variability expected for a group of equivalent subjects, respectively.

σ_{rep} 3-22% [21], σ_{rep} 3-7% [22];
σ_{intra} 1-4% [22], $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2}$ 4-7% [23], $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2}$ 4-8% [5]; $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2}$ 4-12% (closely linked to CRMVB)[24], $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2}$ 6-28% [25], $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2}$ 9-17% (Hippocampus)[26],
σ_{inter} 4-9% [23], $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2 + \sigma_{inter}^2}$ 8-15% [5], $\sqrt{\sigma_{intra}^2 + \sigma_{rep}^2 + \sigma_{inter}^2}$ 7-16% [27].
Above values are for SV studies, similar numbers can be found for CSI data[11;28-30].

Choice of normal control data:

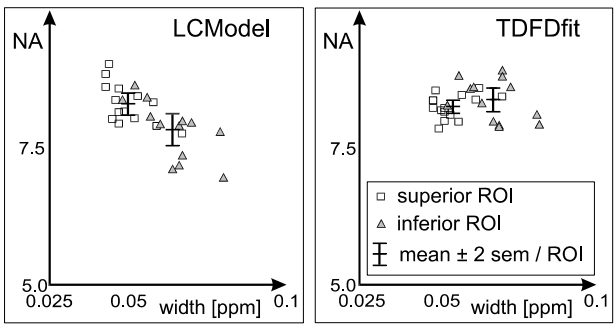

Fig. 4: *Results for fitting with 2 different programs and differing models and fitting constraints for NA_{tot} (NAA + NAAG) as function of linewidth and location for 26 spectra from medial frontal cortex.* [7]

Usually one tries to have a set of high quality spectra obtained from healthy subjects to serve as control data, the rationale being that one does not want to increase variability from the control cohort. This is reasonable as long as the measured quantity does not depend on the quality of the spectra. As mentioned above this is not always the case[5-7]. Fig. 4 contains data obtained at different levels in medial frontal cortex in a group of healthy subjects. Using LCModel the NA content seemed to vary with the position in frontal cortex. However, careful analysis showed that the NA content depended on the spectral resolution which happened to be significantly different for the two locations. Analysis using a different program with different prior knowledge constraints revealed the apparent location dependence to be artifactual[‡]. Since patients are often less compliant or more restless than volunteers, control data can easily be of better quality than patient data. Straightforward comparison may show apparent metabolic differences that can be due to this systematic effect of line width. The example suggests that control data should match patient data in data quality, unless line width and/or SNR is taken as a covariate. Similar caveats apply to results recorded with spectral editing methods. Additionally, one should be aware that normal ranges, i.e. in particular the cohort SD, obtained from high quality control spectra may not be suited to judge normality of patient data deduced from spectra of inferior quality.

Criteria to define abnormality: A single metabolite level in a single spectrum from a single subject can be considered abnormal, if it lies outside the normal range defined by the mean ± 2 SD of the control cohort. Additionally, SNR and FWHM should also fall within normal limits. The control values must originate from truly comparable exams (ROI size, acquisition parameters, location, subject age, data quality). Given the reproducibility found in the literature, deviations from the norm must usually be at least 15%, possibly much more, depending on which metabolite is considered, one's own reproducibility values, and ROI size. Smaller changes may be confirmed in repeated studies, but should still be judged against incidental individuality (σ_{inter}). In CSI studies, detection of abnormalities can be based on similar principles, but CSI also offers additional options. One possibility is to compare intraindividually with the contralateral side, others are to compare correlations with tissue content[31].

[‡] This example does by no means try to convey the claim of a general inferior performance of LCModel. In other circumstances TDFDfit showed similar behavior. The example only aims to show that spectral quality can influence numeric results.

Reference List

1. Kreis R. Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts. NMR Biomed. 2004;17:361-381.
2. Cavassila S, Deval S, Huegen C, van Ormondt D, Graveron-Demilly D. Cramer-Rao bounds: an evaluation tool for quantitation. NMR Biomed. 2001;14:278-283.
3. Kreis R, Hofmann L, Kuhlmann B, Boesch C, Bossi E, Hüppi PS. Brain metabolite composition during early human brain development as measured by quantitative in vivo 1H magnetic resonance spectroscopy. Magn Reson Med. 2002;48:949-958.
4. Ebel A, Soher BJ, Maudsley AA. Assessment of 3D proton MR echo-planar spectroscopic imaging using automated spectral analysis. Magn Reson Med. 2001;46:1072-1078.
5. Schirmer T, Auer DP. On the reliability of quantitative clinical magnetic resonance spectroscopy of the human brain. NMR Biomed. 2000;13:28-36.
6. Tkac, I., Ugurbil, K., and Gruetter, R. On the quantification of low concentration metabolites by 1H NMR spectroscopy in the human brain at 7 Tesla. 10th Meeting of the ISMRM 2002, 528.
7. Kreis, R. and Boesch, C. Bad spectra can be better than good spectra. 11th Meeting of the ISMRM 2003, 264.
8. Ratiney H, Sdika M, Coenradie Y, Cavassila S, van Ormondt D, Graveron-Demilly D. Time-domain semi-parametric estimation based on a metabolite basis set. NMR Biomed. 2005;18:1-13.
9. Kreis R. Quantitative localized ^{1}H-MR spectroscopy for clinical use. Prog NMR Spectroscopy. 1997;31:155-195.
10. Young K, Khetselius D, Soher BJ, Maudsley AA. Confidence images for MR spectroscopic imaging. Magn Reson Med. 2000;44:537-545.
11. McLean MA, Woermann FG, Barker GJ, Duncan JS. Quantitative analysis of short echo time (1)H-MRSI of cerebral gray and white matter. Magn Reson Med. 2000;44:401-411.
12. Felblinger J, Kreis R, Boesch C. Effects of physiologic motion of the brain upon quantitative 1H-MRS: Analysis and correction by retro-gating. NMR Biomed. 1998;11:107-114.
13. Haupt CI, Kiefer AP, Maudsley AA. In-plane motion correction for MR spectroscopic imaging. Magn Reson Med. 1998;39:749-753.
14. Simmons A, Smail M, Moore E, Williams SCR. Serial precision of metabolite peak area ratios and water referenced metabolite peak areas in proton MR spectroscopy of the human brain. Magn Reson Imaging. 1998;16:319-330.
15. Keevil SF, Barbiroli B, Brooks JCW, Cady EB, Canese R, et al. Absolute metabolite quantification by in vivo NMR spectroscopy: II. A multicentre trial of protocols for in vivo localised proton studies of human brain. Magn Reson Imaging. 1998;16:1093-1106.
16. Hajek M, Burian M, Dezortova M. Application of LCModel for quality control and quantitative in vivo 1H MR spectroscopy by short echo time STEAM sequence. MAGMA. 2000;10:6-17.
17. Ebel A, Maudsley AA. Detection and correction of frequency instabilities for volumetric 1H echo-planar spectroscopic imaging. Magn Reson Med. 2005;53:465-469.
18. Ryner LN, Ke Y, Thomas MA. Flip angle effects in STEAM and PRESS-optimized versus sinc RF pulses. J Magn Reson. 1998;131:118-125.
19. Keevil SF, Newbold MC. The performance of volume selection sequences for in vivo NMR spectroscopy: implications for quantitative MRS. Magn Reson Imaging. 2001;19:1217-1226.
20. Hennig J. The application of phase rotation for localized in vivo proton spectroscopy with short echo times. J Magn Reson. 1992;96:40-49.
21. Bartha R, Drost DJ, Menon RS, Williamson PC. Comparison of the quantification precision of human short echo time ^{1}H spectroscopy at 1.5 and 4.0 Tesla. Magn Reson Med. 2000;44:185-192.
22. Brooks WM, Friedman SD, Stidley CA. Reproducibility of 1H-MRS in vivo. Magn Reson Med. 1999;41:193-197.
23. Kreis, R., Fusch, C., Maloca, P., Felblinger, J., and Boesch, C. Supposed pathology may be individuality: interindividual and regional differences of brain metabolite concentrations determined by 1H MRS. 2nd Meeting of the Society of Magnetic Resonance, 1994, 45.
24. Geurts JJ, Barkhof F, Castelijns JA, Uitdehaag BM, Polman CH, Pouwels PJ. Quantitative 1H-MRS of healthy human cortex, hippocampus, and thalamus: metabolite concentrations, quantification precision, and reproducibility. J Magn Reson Imaging. 2004;20:366-371.

25. Wellard RM, Briellmann RS, Jennings C, Jackson GD. Physiologic variability of single-voxel proton MR spectroscopic measurements at 3T. AJNR Am J Neuroradiol. 2005;26:585-590.
26. Hammen T, Stadlbauer A, Tomandl B, Ganslandt O, Pauli E, Huk W, Neundorfer B, Stefan H. Short TE single-voxel 1H-MR spectroscopy of hippocampal structures in healthy adults at 1.5 Tesla--how reproducible are the results? NMR Biomed. 2005;18:195-201.
27. Hofmann L, Slotboom J, Jung B, Maloca P, Boesch C, Kreis R. Quantitative 1H-magnetic resonance spectroscopy of human brain: Influence of composition and parameterization of the basis set in linear combination model fitting. Magn Reson Med. 2002;48:440-453.
28. Wiedermann D, Schuff N, Matson GB, Soher BJ, Du AT, Maudsley AA, Weiner MW. Short echo time multislice proton magnetic resonance spectroscopic imaging in human brain: metabolite distributions and reliability. Magn Reson Imaging. 2001;19:1073-1080.
29. Chard DT, McLean MA, Parker GJ, MacManus DG, Miller DH. Reproducibility of in vivo metabolite quantification with proton magnetic resonance spectroscopic imaging. J Magn Reson Imaging. 2002;15:219-225.
30. Li BS, Babb JS, Soher BJ, Maudsley AA, Gonen O. Reproducibility of 3D proton spectroscopy in the human brain. Magn Reson Med. 2002;47:439-446.
31. Hetherington HP, Pan JW, Mason GF, Adams D, Vaughn MJ, Twieg DB, Pohost GM. Quantitative 1H spectroscopic imaging of human brain at 4.1 T using image segmentation. Magn Reson Med. 1996;36:21-29.
32. Provencher SW. Estimation of metabolite concentration from localized in vivo proton NMR spectra. Magn Reson Med. 1993;30:672-679.

Acknowledgement: *Supported by the Swiss National Foundation (31-103938)*